The Annals of Mathematical Statistics
1971, Vol. 42, No. 6, 1897–1908

PROCEDURES FOR REACTING TO A CHANGE IN DISTRIBUTION

BY G. LORDEN

California Institute of Technology

0. Summary. A problem of optimal stopping is formulated and simple rules are proposed which are asymptotically optimal in an appropriate sense. The problem is of central importance in quality control and also has applications in reliability theory and other areas.

1. Introduction. Suppose $X_1, X_2, \cdots$ are independent random variables observed sequentially and $X_1, \cdots, X_{m-1}$ have distribution function F_0 while $X_m, X_{m+1}, \cdots$ have distribution function $F_1 \neq F_0$. Both F_0 and F_1 are known but m is unknown and some action should be taken after $X_m, X_{m+1}, \cdots$ begin appearing, the sooner the better. As a possible procedure we consider every (nonrandomized) stopping variable N with respect to the observed sequence. Thus the event $\{N = n\}$, which denotes stopping to take action after observing $X_1, \cdots, X_n$, is determined by $X_1, \cdots, X_n$ (i.e. belongs to the sigma-field generated by $X_1, \cdots, X_n$). For $m = 1, 2, \cdots$, let P_m denote the distribution of the sequence $X_1, X_2, \cdots$ under which X_m is the first term with distribution function F_1. If P_m is the true distribution, then in the event that $N \geqq m$ it is desired that the conditional expectation of $N-m$ should be small. Letting E_m denote expectation under P_m, we therefore define

$$\bar{E}_1 N = \sup_{m \geqq 1} \operatorname{ess\,sup} E_m[(N-m+1)^+ \mid X_1, \cdots, X_{m-1}] \tag{1}$$

to serve as a "minimax" type of criterion for quickness of reaction to a change. Let P_0 denote the distribution under which $X_1, X_2, \cdots$ (independent) have distribution function F_0. It is assumed that the desire for small $\bar{E}_1 N$ is offset by the requirement that the frequency of "false reactions" be controlled by a condition of the form $E_0 N \geqq \gamma$, for a prescribed $\gamma > 0$. In other words, subject to $E_0 N \geqq \gamma$, we seek to minimize $\bar{E}_1 N$, which is the smallest A such that for $m = 1, 2, \cdots$

$$E_m[N-(m-1) \mid X_1 = x_1, \cdots, X_{m-1} = x_{m-1}] \leqq A$$

for almost every (P_0) point $(x_1, \cdots, x_{m-1})$ in $\{N \geqq m\}$. Thus $\bar{E}_1 N$ is the smallest bound on the average number of differently distributed X's observed before reacting, guaranteed regardless of the behavior of the X's before the change.

In Section 2 the smallest possible $\bar{E}_1 N$ is determined asymptotically as $\gamma \to \infty$ and shown to be attained asymptotically by the following procedure due to Page (1954): stop the first time

$$T_n = \sum_{i=1}^{n} \log \frac{f_1(X_i)}{f_0(X_i)} - \min_{k \leqq n} \sum_{i=1}^{k} \log \frac{f_1(X_i)}{f_0(X_i)} > \gamma, \tag{2}$$

Received July 27, 1970.

where f_0, f_1 are densities of F_0, F_1 with respect to a suitable measure (e.g. $\frac{1}{2}(F_0+F_1)$). For computational purposes the obvious recursion formula

$$T_n = \left(T_{n-1} + \log\frac{f_1(X_n)}{f_0(X_n)}\right)^+ \tag{3}$$

is useful. Page also pointed out that the procedure in (2) is equivalent to performing a sequential probability ratio test (SPRT) of f_0 vs. f_1 with log-boundaries 0 and γ and repeating the test on successive new observations until a decision in favor of f_1 is reached. It follows by a routine application of Wald's equation that for this procedure

$$E_0 N = \alpha^{-1} E_0 N^*, \tag{4}$$

where N^* is the time for a single SPRT to stop and $\alpha \doteq P_0$ (decide f_1), so that the expected number of tests required is α^{-1}. Similarly,

$$E_1 N = (1-\beta)^{-1} E_1 N^*, \tag{5}$$

where $\beta = P_1$ (decide f_0), and it is clear from the definition of T_n that $\bar{E}_1 N = E_1 N$. Page obtained these formulas for $E_0 N$ and $E_1 N$, which he called the average run length (A.R.L.) under P_0 and P_1, respectively. Although the A.R.L., $E_1 N$, is equal to $\bar{E}_1 N$ for Page's procedure, this need not be the case for alternative procedures, and $E_1 N$ is clearly an inadequate criterion for performance. (Note that an unrepeated SPRT with log-boundaries $-\infty$ and 0 yields $E_0 N = \infty$, in fact $P_0(N = \infty) > 0$, and yet $E_1 N$ will be quite small; alas, $\bar{E}_1 N = \infty$).

In both a practical and a theoretical sense, the problem becomes more interesting if we replace F_1 by a dominated family of distributions $\{F_\theta, \theta \in \Theta\}$ with θ unknown and try to achieve small $\bar{E}_\theta N$ (defined like $\bar{E}_1 N$) for each θ, subject to $E_0 N \geqq \gamma$. It is to be expected that one cannot simultaneously minimize for all θ; however, the results of Section 2 indicate that one can simultaneously minimize for each θ asymptotically as $\gamma \to 0$ for a wide class of problems. These results will be established by exploiting the connection between the present problem and one-sided sequential testing. To see this connection, consider the following alternative description of Page's procedure: stop the first time

$$\max_{k \leqq n} \sum_{i=k}^{n} \log\frac{f_1(X_i)}{f_0(X_i)} > \gamma. \tag{6}$$

This can be regarded as a "maximum likelihood" treatment of the unknown change point, i.e. stop when *for some k* the observations $X_k, \cdots, X_n$ are "significant." In this case, significance is measured by a one-sided SPRT with right-hand log-boundary γ. To handle composite $\{F_\theta\}$ one can simply use instead a one-sided sequential test of F_0 vs. $\{F_\theta\}$ and apply it to $X_k, X_{k+1}, \cdots$ for each $k = 1, 2, \cdots$, stopping the first time one of these tests says stop. This approach is shown to work in Section 2 where the following main result is obtained.

THEOREM 1. *Suppose there exists a class of "one-sided tests" (i.e. extended stopping variables)* $\{N(\alpha), 0 < \alpha < 1\}$ *such that*

$$P_0(N(\alpha) < \infty) \leqq \alpha \qquad \textit{for all } \alpha,$$

and *for all* $\theta \in \Theta$

$$E_\theta N(\alpha) \sim \frac{|\log \alpha|}{I(\theta)} \qquad \textit{as } \alpha \to 0$$

where

$$I(\theta) = E_\theta \log \frac{f_\theta(X)}{f_0(X)} < \infty \qquad \textit{for all } \theta.$$

For $\gamma > 1$ *let* $\alpha = \gamma^{-1}$ *and define* $N^*(\gamma) = \min_{k \geqq 1} \{N_k(\alpha) + k - 1\}$, *where* $N_k(\alpha)$ *is* $N(\alpha)$ *applied to* $X_k, X_{k+1}, \cdots$. *Then* $N^*(\gamma)$ *is a stopping variable,*

$$E_0 N^*(\gamma) \geqq \gamma \qquad \textit{for all } \gamma, \tag{7}$$

and for all $\theta \in \Theta$ $\{N^*(\gamma), \gamma > 1\}$ *minimizes* $\bar{E}_\theta N^*(\gamma)$ *asymptotically subject to* (7), *by virtue of the relation*

$$\bar{E}_\theta N^*(\gamma) \sim \frac{\log \gamma}{I(\theta)} \qquad \textit{as } \gamma \to \infty. \tag{8}$$

The fact that this last relation characterizes asymptotic optimality is established by Theorem 3 of Section 2. The existence of asymptotically optimal one-sided tests is guaranteed under fairly mild conditions by Lemmas 1 and 2 of (Kiefer and Sacks, 1963). The tests are one-sided SPRT's of F_0 against a (fully-supported) mixture of F_θ's, an approach described in (Wald, 1947). Unfortunately, these tests are difficult to perform, in general, because a nontrivial integration is necessary at each stage to determine the likelihood ratio. The resulting reaction procedures require computation of n such likelihood ratios after the nth observation. In order to derive useful asymptotically optimal procedures, Section 3 is restricted to the case where F_0 and the F_θ's belong to a Koopman-Darmois family. In this context, asymptotic optimality is attained by simple one-sided test procedures based on the maximum likelihood ratio (in θ). These give rise to explicit stopping boundaries like those in Schwarz's (1962) work on asymptotic shapes of Bayes sequential testing regions. The use of these procedures is fully described in Section 3, which is independent of Section 2. It should be noted that recent work by Robbins and Siegmund (to appear) shows that the approach based on mixtures can indeed be used to derive explicit procedures for some problems, particularly those involving normal distributions. For further discussion and comparisons between this method and the maximum likelihood approach, the reader is referred to [5]. Some applications of maximum likelihood reaction procedures are given in Section 4.

Besides Page's (1954) work, the following are a few of the many other approaches to the formulation and solution of problems involving the detection of changes in distribution. The classical Shewhart (1931) control charts using "3σ limits" have been widely used in quality control applications. An interesting Bayesian formulation utilizing geometric prior distributions for the unknown change point was studied by Girshick and Rubin (1952). Chernoff and Zacks (1964) and later Zacks and Kander (1966) considered a sort of "retrospective" formulation in which one looks at a fixed sample $X_1, \cdots, X_n$ and attempts to determine whether and where a change has occurred. They gave no optimality results, but did compare their procedures numerically with those of Page, which were also proposed for problems of this fixed-sample type (1955). A variety of practical procedures have been used and numerical comparisons of their performance are given in (Roberts, 1966), using the "worst case" to measure quickness of reaction, which is tantamount to adopting the minimax-type criterion, $\bar{E}_1 N$, defined in (1).

2. Asymptotic theory. The first result establishes bounds on the performance of reaction procedures constructed from one-sided tests.

THEOREM 2. *Let N be an extended stopping variable with respect to $X_1, X_2, \cdots$ such that*

$$P_0(N < \infty) \leqq \alpha. \tag{9}$$

For $k = 1, 2, \cdots$ let N_k denote the stopping variable obtained by applying N to $X_k, X_{k+1}, \cdots$, and define

$$N^* = \min\{N_k + k - 1 \mid k = 1, 2, \cdots\}.$$

Then N^* is an extended stopping variable,

$$E_0 N^* \geqq 1/\alpha, \tag{10}$$

and for any alternative distribution, F_1,

$$\bar{E}_1 N^* \leqq E_1 N. \tag{11}$$

PROOF. N^* is an extended stopping variable since the event $\{N^* \leqq n\}$ is the union of $\{N_1 \leqq n\}, \{N_2 \leqq n-1\}, \cdots, \{N_n \leqq 1\}$, all of which are evidently determined by $X_1, \cdots, X_n$. For $m = 1, 2, \cdots$

$$E_m[(N^* - m + 1)^+ \mid X_1, \cdots, X_{m-1}] \leqq E_m[N_m \mid X_1, \cdots, X_{m-1}] = E_m N_m = E_1 N,$$

and (11) follows by definition of $\bar{E}_1 N^*$.

To prove (10), define

$$\begin{aligned} \xi_k &= 1 \quad \text{if} \quad N_k < \infty, \\ &= 0 \quad \text{if} \quad N_k = \infty, \qquad k = 1, 2, \cdots. \end{aligned}$$

Since the ergodic hypothesis is true for the i.i.d. sequence $X_1, X_2, \cdots$ (Loéve, (1963)),

$$\lim_{n\to\infty} n^{-1} \textstyle\sum_{k=1}^n \xi_k = E_0 \xi_1 = P_0(N_1 < \infty) \leqq \alpha \text{ a.s.} (P_0). \tag{12}$$

Assume that $E_0 N^* < \infty$ (otherwise (10) is trivial). Let $N_0{}^* = 0$ and define $N_1{}^* < N_2{}^* < \cdots$ recursively as follows. If $N_{m-1}^* = n$, then for each $r = 1, 2, \cdots$ apply N to $X_{n+r}, X_{n+r+1}, \cdots$ and let $N_m{}^*$ be the first time stopping occurs for some r. Then $N_1{}^* = N^*$ and $N_1{}^*, N_2{}^* - N_1{}^*, N_3{}^* - N_2{}^*, \cdots$ are independent and identically distributed.

Clearly

$$\xi_{N_m^*+1} + \cdots + \xi_{N_{m+1}^*} \geqq 1 \qquad \text{for } m = 0, 1, \cdots,$$

since $\xi_{N_m^*+r} = 1$ for some r causing the stop at N_{m+1}^*. Hence, $\xi_1 + \cdots + \xi_{N_m^*} \geqq m$ for $m = 0, 1, \cdots$, so that

$$\frac{\xi_1 + \cdots + \xi_{N_m^*}}{N_m^*} \geqq \frac{m}{N_m^*}. \tag{13}$$

As $m \to \infty$, the right-hand side of (13) approaches $(E_0 N^*)^{-1}$ by the strong law of large numbers and the left-hand side tends to a limit $\leqq \alpha$ by (12), proving (10).

REMARK. By a similar renewal argument it can be shown that if $N_1, \cdots, N_k$ are stopping variables and $N = \min(N_1, \cdots, N_k)$, then $(E_0 N)^{-1} \leqq (E_0 N_1)^{-1} + \cdots + (E_0 N_k)^{-1}$. For example, if $E_0 N_i \geqq k\gamma$, $i = 1, \cdots, k$, then $E_0 N \geqq \gamma$.

If N is the stopping variable of a one-sided SPRT of F_0 vs. F_1 with likelihood-ratio boundary $1/\alpha$, then by well-known results (Wald (1947)) $P_0(N < \infty) \leqq \alpha$ and we have $E_1 N \sim |\log \alpha|/I_1$, as $\alpha \to 0$, where I_1 is the information number, $E_1 \log(f_1(X)/f_0(X))$. Applying Theorem 2, we obtain an N^* (Page's procedure) satisfying $E_0 N^* \geqq \alpha^{-1}$ and $\bar{E}_1 N^*$ asymptotically at most $|\log \alpha|/I_1$ as $\alpha \to 0$. Theorem 3 establishes that this is asymptotically the best one can do.

THEOREM 3. *Let $n(\gamma)$ be the infimum of $\bar{E}_1 N$ as N ranges over the class of extended stopping variables satisfying $E_0 N \geqq \gamma$. If $I_1 = E_1 \log[f_1(X)/f_0(X)] < \infty$, then*

$$n(\gamma) \sim \frac{\log \gamma}{I_1} \qquad \text{as } \gamma \to \infty. \tag{14}$$

The proof uses the following

THEOREM (WALD). *If N is the sample size of a test of f_0 against f_1 with error probabilities α, β* respectively, then

$$I_1 E_1 N \geqq (1-\beta)\log\frac{1-\beta}{\alpha} + \beta \log\frac{\beta}{1-\alpha} \geqq (1-\beta)|\log\alpha| - \log 2,$$

where $I_1 = E_1 \log[f_1(X)/f_0(X)]$.

PROOF. The first inequality appears in (Wald, 1947, page 197). The second inequality follows from the fact that $\beta \log(1-\alpha)^{-1}$ is nonnegative and $(1-\beta)\log(1-\beta) + \beta\log\beta$ attains minimum value $-\log 2$ when $\beta = \frac{1}{2}$.

PROOF OF THEOREM 3. By virtue of the result just established for Page's procedure, it only remains to show that $n(\gamma)$ is asymptotically no smaller than the right-hand member of (14). It will suffice to show that for every ε in (0, 1) there is a $C(\varepsilon) < \infty$ such that for all stopping variables N

$$I_1 \bar{E}_1 N \geqq (1-\varepsilon)\log E_0 N - C(\varepsilon). \tag{15}$$

Fix ε and define stopping variables $T_0 \equiv 0 < T_1 < T_2 < \cdots$ as follows: $T_{i+1} (i = 0, 1, \cdots)$ is the smallest n (or ∞ if there is no n) such that $n > T_i$ and

$$f_1(X_{T_i+1}) \cdots f_1(X_n) \leqq \varepsilon f_0(X_{T_i+1}) \cdots f_0(X_n). \tag{16}$$

By the standard argument used to estimate error probabilities of sequential probability ratio tests (Wald, 1947), $P_1(T_1 < \infty) \leqq \varepsilon$, and the same argument is easily modified to yield

$$P_{k+1}(T_r < \infty \mid D_{rk}) \leqq \varepsilon \quad \text{provided} \quad P_1(D_{rk}) > 0, \tag{17}$$

where $D_{rk} = \{T_{r-1} = k < N\}$, which depends only on $X_1, \cdots, X_k$.

Consider all those subsets D_{rk} for which $P_0(D_{rk}) > 0$, and hence also $P_{k+1}(D_{rk}) > 0$ because P_{k+1} gives the same distribution of $X_1, \cdots, X_k$ as does P_0. On the subset D_{rk}, N and T_r determine the following sequential test based on $X_{k+1}, X_{k+2}, \cdots$: stop at $\min(N, T_r)$ and

$$\begin{aligned} &\text{decide} \quad P_{k+1} \quad \text{is true if } N \leqq T_r; \\ &\text{decide} \quad P_0 \qquad \text{is true if } N > T_r. \end{aligned}$$

The number of observations taken is $\min(N, T_r) - k$, whose (conditional) expectation under $P_{k+1}(\cdot \mid D_{rk})$ is at most $\bar{E}_1 N$. (D_{rk} belongs to the σ-field of events determined by $X_1, \cdots, X_k$.) Wald's theorem applies with $\alpha = P_0(N \leqq T_r \mid D_{rk})$ and $1-\beta = P_{k+1}(N \leqq T_r \mid D_{rk})$, so that we have

$$\begin{aligned} I_1 \bar{E}_1 N &\geqq P_{k+1}(N \leqq T_r \mid D_{rk}) |\log P_0(N \leqq T_r \mid D_{rk})| - \log 2 \\ &\geqq (1-\varepsilon) |\log P_0(N \leqq T_r \mid D_{rk})| - \log 2, \end{aligned} \tag{18}$$

the latter inequality following from

$$P_{k+1}(N \leqq T_r \mid D_{rk}) \geqq P_{k+1}(T_r = \infty \mid D_{rk})$$

and (17).

Let R be the smallest $r \geqq 1$ (or ∞ if there is no r) such that $T_r > N$. If $P_0(R \geqq r) > 0$, then $P_0(R < r+1) \mid R \geqq r)$ is well defined and evidently equals $P_0(N \leqq T_r \mid T_{r-1} < N)$, which is an average (over k) of the probabilities $P_0(N \leqq T_r \mid T_{r-1} = k < N)$ satisfying (18). Therefore $P_0(R \geqq r) > 0$ implies

$$I_1 \bar{E}_1 N \geqq (1-\varepsilon) |\log P_0(R < r+1 \mid R \geqq r)| - \log 2. \tag{19}$$

Elementary calculations show that a lower bound of the form

$$P_0(R < r+1 \mid R \geqq r) \geqq Q \quad \text{for} \quad r = 1, 2, \cdots, \text{such that} \quad P_0(R \geqq r) > 0$$

implies

$$P_0(R \geqq r+1) \leqq (1-Q)^r \qquad \text{for } r = 1, 2, \cdots$$

and hence $E_0 R \leqq Q^{-1}$. Thus (19) yields

$$I_1 \bar{E}_1 N \geqq (1-\varepsilon) \log E_0 R - \log 2. \tag{20}$$

When P_0 is true $\{T_i\}$ is a sequence of cumulative sums of independent random variables distributed like T_1, as can be verified from (16). Since $E_0R < \infty$ by (20), Wald's equation

$$E_0 T_R = E_0 R \cdot E_0 T_1$$

holds, and hence

$$\log E_0 N \leqq \log E_0 T_R = \log E_0 R + B(\varepsilon), \tag{21}$$

where $B(\varepsilon) = \log E_0 T_1$, which is finite for all ε and does not depend on N. Relation (15) follows at once from (20) and (21), and the proof is complete.

PROOF OF THEOREM 1. For each $\theta \in \Theta$ relations (7) and (8) follow from the hypotheses about $N(\alpha)$ by Theorem 2. The fact that these relations characterize asymptotic optimality is the content of Theorem 3.

3. Reaction procedures for Koopman-Darmois families. Suppose that the F_θ's are members of a Koopman-Darmois family of distributions, i.e.

$$dF_\theta(x) = \exp(\theta T(x) - b(\theta))\, d\mu(x), \qquad \theta \in \Theta^*, \tag{22}$$

where μ is a σ-finite measure on the real Borel sets and Θ^* is an interval on the real line. Then $b(\theta)$ is strictly concave upward and infinitely differentiable on Θ^*. We assume that F_0 is also a member of the same family. There is no loss of generality in assuming that F_0 corresponds to $\theta = 0$ (shifting Θ^* if necessary) and that $b(0) = 0$ (by incorporating $\exp(-b(0))$ with μ). Thus μ becomes identified with F_0 and by regarding $T(X_1)$, $T(X_2)$, $\cdots$ rather than the X's as the observations we can rewrite (22) in the simpler form

$$dF_\theta(x) = \exp(\theta x - b(\theta))\, dF_0(x), \qquad \theta \in \Theta^*, \tag{23}$$

where Θ^* contains 0 and $b(0) = 0$. Let $\Theta = \Theta^* - \{0\}$.

In order to obtain asymptotically optimal procedures by application of Theorem 1 we need only determine stopping variables, $N(\alpha)$, such that

$$P_0(N(\alpha) < \infty) \leqq \alpha \qquad \text{for } 0 < \alpha < 1 \tag{24}$$

and

$$E_\theta N(\alpha) \sim \frac{|\log \alpha|}{I(\theta)} \quad \text{as} \quad \alpha \to 0, \qquad \text{for all } \theta \in \Theta. \tag{25}$$

A routine calculation shows that $I(\theta) = \theta b'(\theta) - b(\theta)$. The log-likelihood-ratio of F_θ over F_0 based on $X_1, \cdots, X_n$ equals $\theta S_n - nb(\theta)$, where $S_n = X_1 + \cdots + X_n$. The procedures $N(\alpha)$ will be given by rules of the following form: stop at the first n such that

$$\sup_{|\theta| \geqq \theta_1(\alpha)} (\theta S_n - nb(\theta)) > h(\alpha) > 0, \tag{26}$$

where $\theta_1(\alpha) \downarrow 0$ and $h(\alpha) \to \infty$ as specified later. Note that if 0 is an endpoint of Θ^*, which may be desired in some applications, then $|\theta| \geqq \theta_1(\alpha)$ in (26) is still correct, it being understood that θ ranges over Θ^* only (which need not be the full natural parameter space of the family).

It is clear from (26) that $N(\alpha)$ is obtained by stopping the first time that a one-sided SPRT of F_0 vs. F_θ with log-boundary $h(\alpha)$ says stop for some θ with $|\theta| \geqq \theta_1(\alpha)$. Since $\theta_1(\alpha) \downarrow 0$ as $\alpha \to 0$, any fixed $\theta \neq 0$ is included for sufficiently small α and hence $E_0 N(\alpha)$ is at most the expected time for a one-sided SPRT of F_0 vs. F_θ. The latter is asymptotically $h(\alpha)/I(\theta)$ as $h(\alpha) \to \infty$, so that

$$\limsup_{\alpha \downarrow 0} \frac{E_\theta N(\alpha)}{h(\alpha)} \leqq \frac{1}{I(\theta)}. \tag{27}$$

It evidently suffices for (25) to choose $h(\alpha) \sim |\log \alpha|$. To see how to satisfy (24), we next consider the problem of estimating the "error probability," $P_0(N(\alpha) < \infty)$.

Begin by rewriting (26) in the equivalent form

$$S_n > \inf_{\theta \geqq \theta_1(\alpha)} \left\{\frac{h(\alpha)}{\theta} + n\frac{b(\theta)}{\theta}\right\} \quad \text{or} \tag{28}$$

$$S_n < \sup_{\theta \leqq -\theta_1(\alpha)} \left\{\frac{h(\alpha)}{\theta} + n\frac{b(\theta)}{\theta}\right\}.$$

Fix n. If the infimum in (28) is attained at $\theta = \theta_n$ (say), then the probability that the first inequality holds is at most $\exp(-h(\alpha))$ because the latter is the standard upper bound on the probability of ever terminating the one-sided SPRT of $\theta = 0$ against $\theta = \theta_n$. The same upper bound applies even if the infimum is not attained, since we can apply the same argument to a sequence of θ's along which the infimum is approached. Similar reasoning applies to the second inequality in (28), leading to the conclusion that

$$P_0\{N(\alpha) = n\} \leqq 2\exp(-h(\alpha)) \qquad \text{for } n = 1, 2, \cdots. \tag{29}$$

By differentiating the right-hand member of the first inequality in (28), one finds that the infimum is attained either at θ_1, or as θ approaches an endpoint of Θ^*, or for θ satisfying $nI(\theta) = h(\alpha)$. Since $I(\theta)$ increases with θ, the first of these is the case unless $n < h(\alpha)I(\theta_1)^{-1}$. The same argument used for (29) now shows that

$$P_0\left\{\left[\frac{h(\alpha)}{\min(I(\theta_1), I(-\theta_1))}\right] < N(\alpha) < \infty\right\} \leqq 2\exp(-h(\alpha)), \tag{30}$$

where $[x]$ denotes the largest integer $\leqq x$. Using (29) for n not included in the estimate (30), we obtain

$$P_0\{N(\alpha) < \infty\} \leqq 2\exp(-h(\alpha))\left\{\frac{h(\alpha)}{\min(I(\theta_1), I(-\theta_1))} + 1\right\}. \tag{31}$$

It is known (e.g. Wong, (1968)) that the error probability is of smaller order of magnitude than the right-hand side of (31). However, the latter bound is simple and explicit, and suffices for the present application.

We now indicate how to choose $\theta_1(\alpha)$ and $h(\alpha)$ and verify (24) and (25). First set $\theta_1(\alpha) = |\log \alpha|^{-1}$ or $\frac{1}{3}$ times the length of Θ^*, whichever is smaller. Then choose $h(\alpha) \geqq 1$ as small as possible so that the right-hand side of (31), which is decreasing in $h(\alpha)$, is at most α. This last choice makes (24) an immediate consequence of (31), and it remains only to show that the $h(\alpha)$ so defined is asymptotic to $|\log \alpha|$, which suffices for (25) (see the remark following (27)). By expanding $b(\theta)$ and $b'(\theta)$ in Taylor series about zero, it is seen that $I(\theta) = \theta b'(\theta) - b(\theta) \sim C\theta^2$ as $\theta \to 0$, where $C = \frac{1}{2}b''(0) > 0$. Therefore, the above choices of θ_1 and $h(\alpha)$ imply

$$(32) \qquad 2h(\alpha)\exp(-h(\alpha)) \sim C\alpha|\log\alpha|^{-2} \qquad \text{as } \alpha \to 0.$$

Substituting $(1+\varepsilon)$ $|\log \alpha|$ for $h(\alpha)$ makes the left member of (32) asymptotically too small if $\varepsilon > 0$ and too large if $\varepsilon < 0$. Therefore, $h(\alpha) \sim |\log \alpha|$, as required.

In practice, there are two useful schemes for applying Page's procedure. One is the graphical method of plotting S_n versus n on a cusum (for "cumulative sum") chart. A straight edge moved along and placed at the proper angle and distance from the latest point can be used with the rule "stop if a previously plotted point lies on the opposite side of the straight line." Thus a run of length k is "significant" whenever its sum exceeds $A+Bk$, where A and B are computed to give the desired likelihood-ratio boundary for a 1-sided SPRT. In effect, after the nth observation one performs an SPRT graphically on the reversed sequence $X_n, X_{n-1}, \cdots, X_1$. A considerable simplification is effected by using a parallel boundary through the last plotted point. As soon as any previously plotted point falls across this boundary then that point and all its predecessors can be eliminated from further consideration, since no significant run including these points can occur unless there is also a shorter significant run not involving them. To perform two Page procedures simultaneously, with alternative θ's on either side of the parameter value, one can use a "V-shaped" boundary with the vertex placed at the proper distance in front of the latest point. The maximum likelihood procedures can be used in exactly the same way, replacing the straight-line stopping boundaries by convex boundaries. The values $A+Bk$ for the straight-line boundary are replaced by one of the sequences in (28), i.e.

$$(33) \qquad c_k = \inf_{\theta \geqq \theta_1}\left\{\frac{h(\alpha)}{\theta} + k\frac{b(\theta)}{\theta}\right\} \qquad \text{or}$$

$$c_k = \sup_{\theta \leqq -\theta_1}\left\{\frac{h(\alpha)}{\theta} + k\frac{b(\theta)}{\theta}\right\},$$

depending on whether one is interested in right-hand or left-hand alternatives. If alternatives in both directions are wanted, then one uses a pair of boundaries simultaneously, resulting in a more or less "U-shaped" implement. Several points deserve emphasis. First recall that for $k \geqq h(\alpha)I(\theta_1)^{-1}$ the infimum (resp.supremum) in (33) is attained at θ_1 (resp. $-\theta_1$), so that the boundaries ultimately become straight. Next, it is helpful to recall that the infimum (resp. supremum) for smaller k is attained at θ satisfying $kI(\theta) = h(\alpha)$ (except possibly for some initial values of k where this equation has no solution and the infimum is approached at an endpoint of Θ^*). Finally, it should be noted that, just as for Page's procedure, simplification is achieved by using a straight-line boundary (or two) through the last plotted point, representing an SPRT of $\theta = 0$ vs. $\theta = \theta_1$ with log-boundary 0. As soon as a previously plotted point falls across this line no significant run (for right-hand alternatives) involving that point or its predecessors need be sought.

Carrying out Page's procedure numerically is best handled by the recursive relation (3). To perform a maximum likelihood procedure (for alternative θ's on the right, say), proceed as follows. Compute T_n's recursively by (3) (i.e. perform Page's procedure) with $f_1 = f_{\theta_1}$, the "closest alternative." Stop whenever $T_n \geqq h(\alpha)$ occurs, and whenever $T_n = 0$ one can begin a new cycle, discarding all previous observations and starting fresh on the incoming observations. In addition, each time a new cycle begins compute at each stage $n = 1, 2, \cdots$

$$Q_k^{(n)} = X_n + \cdots + X_{n-k+1}, \qquad k = 1, \cdots, [\min(M, n)]$$

where $M = h(\alpha)I(\theta_1)^{-1}$, stopping at the first n such that

$$Q_k^{(n)} > c_k \qquad \text{for some } k.$$

If $Q_1, \cdots, Q_M$ denote the Q_k's before observing X_n, and $Q_1', \cdots, Q_M'$ denote the current Q_k's , then we have

$$\begin{aligned} Q_1' &= X_n \\ Q_2' &= X_n + Q_1 && \text{(or 0 if } n = 1) \\ &\vdots \\ Q_M' &= X_n + Q_{M-1} && \text{(or 0 if } n < M). \end{aligned}$$

Thus after each observation one can determine by addition a new set of Q's, stopping as soon as some $Q_k > c_k$, and recycling the whole procedure whenever $T_n = 0$. For alternative θ's in both directions one performs simultaneously two procedures as just described, one for each direction of alternatives.

The choice of θ_1 can be made by considering the importance of reacting quickly (or at all) to alternative θ's at various distances from the nominal value. The choice of a critical value ($h(\alpha)$ above) could most effectively be made by comparing the E_0N's attained for various choices with the $\bar{E}_\theta N$'s achieved for alternative θ's. Unfortunately, choosing $h(\alpha)$ to make the right-hand side of (31) equal α is overly pessimistic, in general, because (31) is a crude estimate of the error probability of a one-sided test. Furthermore, the bound (10) in Theorem 2 is also likely to be pessimistic, so that E_0N will tend to be considerably larger than α if the boundary $h(\alpha)$ is chosen as above. It is to be expected that the approximation

$$\bar{E}_\theta N^*(\alpha) \leqq E_\theta N(\alpha) \approx \frac{|\log h(\alpha)|}{I(\theta)}$$

is somewhat better, although it is difficult to estimate its accuracy in general. Perhaps the most satisfactory approach to selection of a critical value is to determine E_0N^* and $\bar{E}_\theta N^*$ for a few θ's by Monte-Carlo methods. The structure of the maximum likelihood procedures indicates that when an alternative θ is true, N^*

should be approximately normally distributed and when $\theta = 0$, N^* should be approximately geometrically distributed (by virtue of the repeated independent "cycles").

In [5], improved methods for approximating error probabilities of the kind in (31) are studied.

4. Examples and applications. A typical application of reaction procedures in quality control is to detect changes in the mean θ of measurements $X_1, X_2, \cdots$ on some manufactured product (or batches), it being assumed that the underlying distribution is normal with variance known, presumably from long-term experience. Assuming the variance is one and the nominal value of θ is zero, we have the information number $I(\theta) = \frac{1}{2}\theta^2$ and for alternative θ's > 0 we have

$$\begin{aligned} c_k &= (2kh(\alpha))^{\frac{1}{2}} && \text{for } k \leqq 2h(\alpha)\theta_1^{-2}; \\ &= \frac{h(\alpha)}{\theta_1} + \tfrac{1}{2}k\theta_1 && \text{for } k > 2h(\alpha)\theta_1^{-2}. \end{aligned}$$

The stopping boundary consists of a piece of a parabola extending into a line tangent to the parabola. It is proved in [5], that in this case, (31) can be improved as follows

$$P_0\{N(\alpha) < \infty\} \leqq 2e^{-h(\alpha)}\left\{1 + \frac{(h(\alpha))^{\frac{1}{2}}\log(2h(\alpha)\theta_1^{-2})}{(4\pi)^{\frac{1}{2}}}\right\}, \tag{34}$$

the factor of two being necessary only when both directions of alternatives are considered.

Another area of application is reliability theory. It is often desired to react to increasing (or decreasing) failure rates, e.g. in detecting the onset of wearout or deterioration of reliability in the course of production. Times between failure are usually assumed to have an exponential distribution or, more generally, a Weibull distribution. Since the Weibull distribution is equivalent to an exponential distribution of a specified power of the failure times, problems of changes in the scale parameter reduce to the exponential case. If the nominal value of the exponential parameter is taken to be 1 (by a scale change, if necessary) then the family of exponential distributions with higher failure rates can be specified conveniently by

$$f_\theta(x) = (1+\theta)\, e^{-\theta x} \qquad (x > 0, \theta > 0),$$

the densities with respect to the nominal distribution. The density f_θ corresponds to a failure rate of $1+\theta$. This leads to

$$c_k = \inf_{\theta \geqq \theta_1}\left\{\frac{h(\alpha)}{\theta} - k\frac{\log(1+\theta)}{\theta}\right\},$$

with stopping as soon as $-S_k > c_k$, i.e. $S_k < -c_k$. (In the Weibull case S_k is the sum of powers of the failure times.) For $k \geqq h(\alpha)\ (\log(1+\theta_1)+(1+\theta_1)^{-1}-1)$ the infimum is attained at θ_1, while for smaller k it turns out that c_k/k is the solution in (0, 1) of $x - \log x = 1 + k^{-1}h(\alpha)$. Of course, the same computations apply to problems of changes in the arrival rates of Poisson processes, which occur in many contexts other than reliability theory.

REFERENCES

[1] CHERNOFF, H. and ZACKS, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *Ann. Math. Statist.* **35** 999–1018.

[2] GIRSHICK, M. A. and RUBIN, H. (1952). A Bayes approach to a quality control model. *Ann. Math. Statist.* **23** 114–125.

[3] KIEFER, J. and SACKS, J. (1963). Asymptotically optimum sequential inference and design. *Ann. Math. Statist.* **34** 705–750.

[4] LOÉVE, M. (1963). *Probability Theory*. Van Nostrand, Princeton.

[5] LORDEN, G. To appear. Open-ended tests for Koopman-Darmois families.

[6] PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115.

[7] PAGE, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika* **42** 523–527.

[8] ROBERTS, S. W. (1966). A comparison of some control chart procedures. *Technometrics* **8** 411–430.

[9] SCHWARZ, G. (1962). Asymptotic shapes of Bayes sequential testing regions. *Ann. Math. Statist.* **33** 224–236.

[10] SHEWHART, W. A. (1931). *The Economic Control of Quality of a Manufactured Product.* Van Nostrand, Princeton.

[11] WALD, A. (1947). *Sequential Analysis.* Wiley, New York.

[12] WONG, S. P. (1968). Asymptotically optimum properties of certain sequential tests. *Ann. Math. Statist.* **39** 1244–1263.

[13] ZACKS, S. and KANDER, Z. (1966). Test procedures for possible changes in parameters of statistical distributions occurring at unknown time points. *Ann. Math. Statist.* **37** 1196–1210.